Laurent.Kling@epfl.ch, EPFL - STI

Accélération de l'innovation
Baisse de la compréhension
Google Hack

La vitesse de l'utilisation des technologies augmente. Sur le graphique ci-dessous, on voit que le temps nécessaire pour qu'une innovation devienne incontournable diminue radicalement. Peut-être que cette accélération entraîne une perte de repères pour les usagers de ces nouvelles technologies. Par exemple, confondre dans ce même graphique Internet, inventé au début des années 1970 avec le Web apparu début 1993.

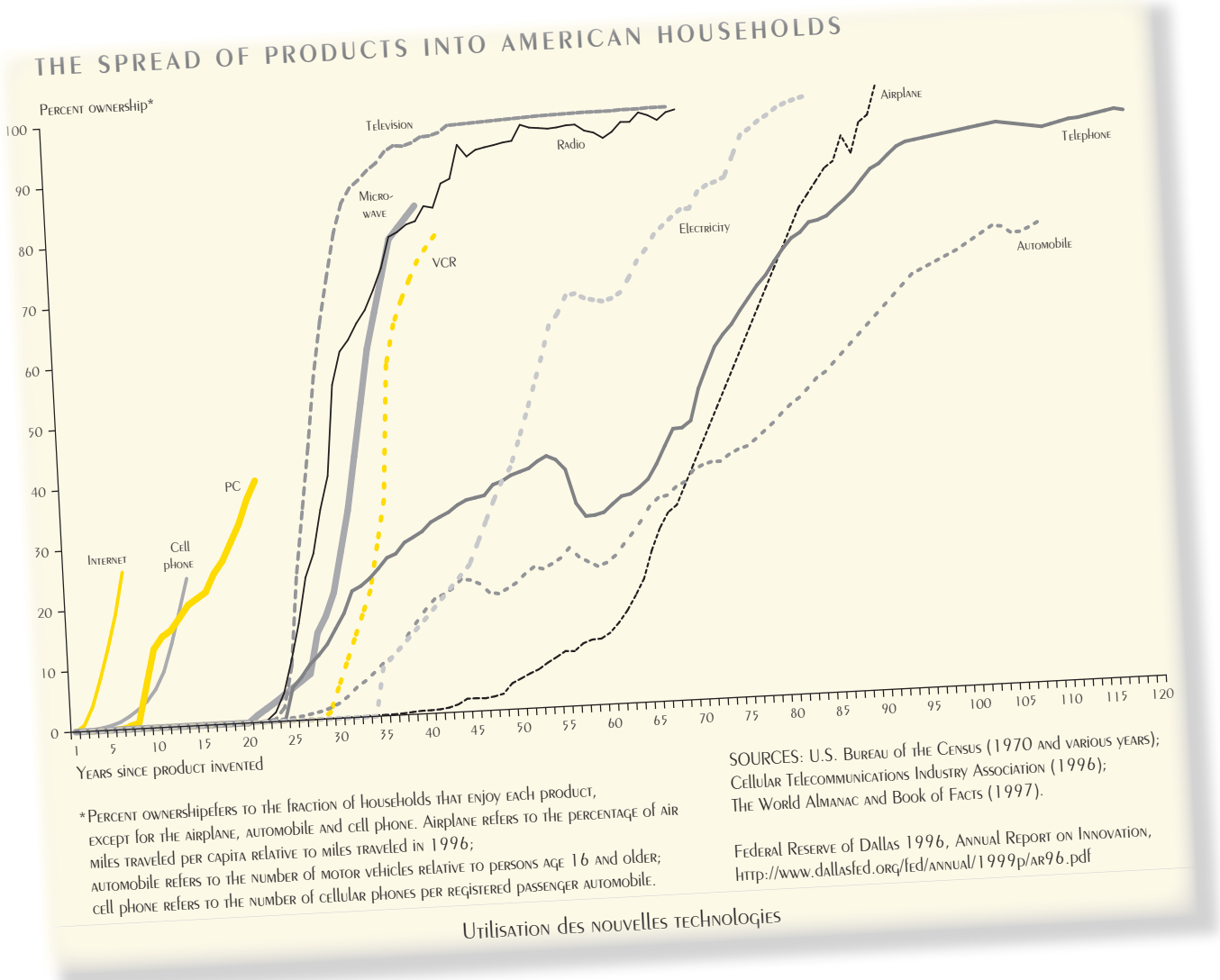
L'utilisateur

L'arrivée du Web a profondément changé les habitudes des usagers dans leurs approches de la sécurité des données. La profusion de sites, blogs et autres outils d'échanges engendre une progression exponentielle des données disponibles.

La recherche sur Internet s'est transformée, de Yahoo (outil de référencement) au moteur de recherche Google, outil quotidien des internautes.

Cette évolution s'est également produite dans l'EPFL, car maintenant, notre moteur de recherche est une boîte noire (peinte en **jaune**), Google Appliance, qui permet de rechercher rapidement l'information.

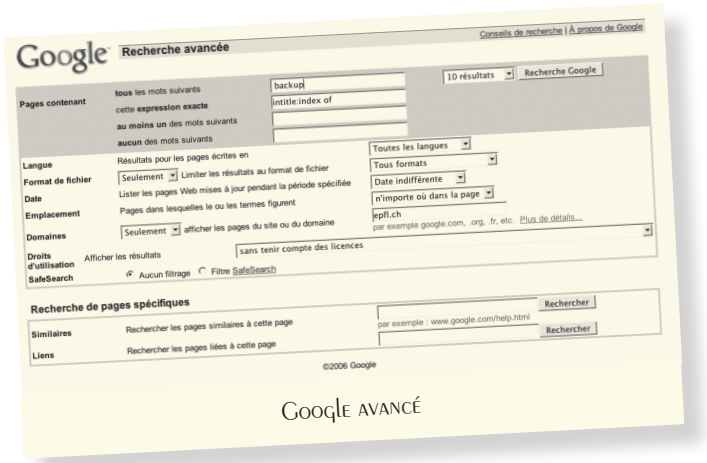
Devant cette avalanche de solutions, l'individu ne peut plus vraiment séparer les contenus privés des contenus publics. En conséquence, la limite entre l'internet privé, Intranet, et l'internet public, Extranet, devient de plus en plus perméable.



Utilisation des nouvelles technologies

Google

Contrairement à un être humain, Google est un processus informatique. Il ne sépare pas le bon grain de l'ivraie dans sa quête frénétique d'information. La lecture de l'article paru dans le FI6/06 (***http://ditwww.epfl.ch/SIC/SA/SPIP/Publications/article.php3?id_article=1107***) permet de comprendre le processus d'indexation utilisé par Google.



Google avancé

Si le résultat fourni par Google est relativement neutre, la clé du processus de recherche consiste à définir, les critères de celle-ci, comme aurait dit La Palice.

La majorité des internautes utilise le moteur de recherche sous son aspect le plus simple, avec parfois une incursion dans le mode de recherche avancée.

Comme souvent, il est parfois plus efficace d'écrire directement sa recherche sous la forme la plus proche du noyau informatique, la ligne de commande.

Abécédaire de commande Google

site:epfl.ch – la commande la plus utile, pour restreindre la recherche à un ensemble DNS, en l'occurrence, l'ensemble des sites du domaine epfl.ch;

intitle:"index of" – une combinaison plus subtile, utiliser le titre de la page Web, puis définir un contenu exact, celui qui est entre guillemets;

filetype:doc – pour restreindre la recherche à un type de document;

cache: – pour rechercher un contenu qui n'existe plus;

intext: – pour rechercher dans le contenu de la page Web;

link: – pour retrouver les pages qui pointent sur une URL

info: – pour afficher les informations que Google connaît.

Google Hack

Par définition, le Web sert à mettre à disposition des informations. Le problème arrive quand l'information n'est plus publique, mais réservée à un groupe restreint d'usagers. Normalement, ces informations devraient résider dans un espace sécurisé avec un accès authentifié. Ainsi, seuls les usagers ayant montré patte blanche peuvent accéder aux données.

À l'EPFL, le NAS avec CIFS et my.epfl.ch, sont deux exemples d'espace de données sécurisé.

Malheureusement, l'apparition du Web a vite entraîné l'idée de conserver des données non publiques dans l'espace de stockage que représente un serveur Web.

L'accès le plus simple sur un serveur Web est l'affichage du contenu d'un dossier.

Le principe de l'utilisation d'un Google Hack est de restreindre la recherche pour obtenir directement les éléments intéressants. Voici cinq exemples de recherche.

Les dossiers privés ou de sauvegarde

Je désire rechercher sur les sites Web référencés pour l'EPFL, les dossiers dont le contenu est visible, mais privé. Formuler en français, cette requête a peu de chance d'être un succès, en mode ligne de commande, le résultat est plus simple:

- rechercher sur le site de l'EPFL: `site:epfl.ch`
- rechercher les dossiers dont le contenu est visible. Cette requête n'existe pas; par la petite porte, on recherche un contenu identifiant cette propriété, le titre de la page Web: `intitle:"index of"`
- rechercher les dossiers privés, pour un public anglophone, il doit comprendre: `private`.

La ligne de commande complète: `intitle:"index of" private site:epfl.ch`.

Une variante sur ce thème, les dossiers de sauvegarde, *backup* pour les anglophones: `intitle:"index of" backup site:epfl.ch`.

Cet exemple est relativement anodin, cela prouve que le contenu d'un serveur Web est réellement disponible pour l'humanité.

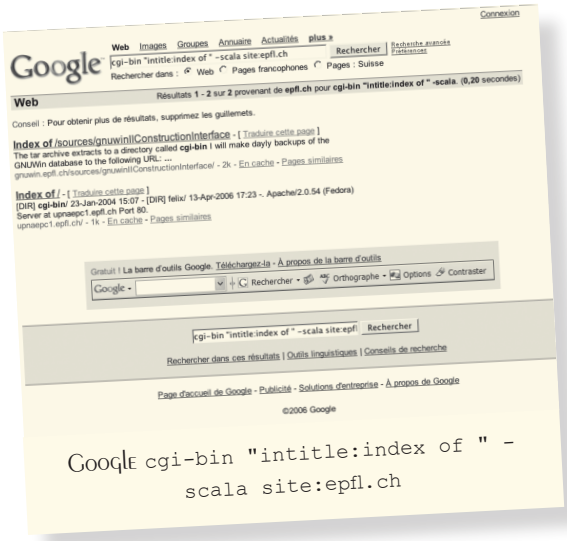
Un site trop visible

Après le succès des premières recherches, un cas plus complexe. Je désire rechercher les dossiers de scripts visibles: `intitle:"index of" cgi-bin site:epfl.ch`.

Google `intitle:"index of" private site:epfl.ch`

Trop de résultats, supprimer les références de "scala..." `cgi-bin "intitle:index of " -scala site:epfl.ch`.

Bien, pas de trou de sécurité, je suis presque tranquille. Étant un lecteur assidu du Flash informatique, j'ai lu qu'une boîte jaune Google est maintenant en service dans l'EPFL, ***http://search.epfl.ch/***.



Google cgi-bin "intitle:index of " -scala site:epfl.ch

gée, ce qui me permet de récupérer certains codes sources. Par acquit de conscience, je me promène sur l'ensemble du site, en étant identifié uniquement par mon adresse IP que je n'ai pas pris la précaution de cacher. Il est utile d'approfondir la notion Proxie Web.

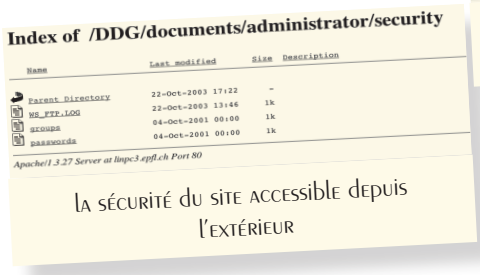
Comme le site est organisé logiquement, je me retrouve dans le dossier: `/DDG/documents/administrator/security/`, et j'obtiens la chaîne de hachage d'un mot de passe administrateur. Évidemment, cette situation est corrigée, mais elle démontre le risque potentiel de mettre un contenu sur le Web!

Google: "Powered by Gallery" site:epfl.ch

Essayons la même recherche en omettant le site, car je suppute qu'uniquement le site EPFL est indexé: `intitle:"index of " cgi-bin`.

Bingo, un site est accessible! `/DDG/cgi-bin/cgi-bin-administrator/` *Nota bene*, ces recherches sont réalisées depuis une connexion extérieure à l'EPFL sans VPN!

Heureusement, le contenu des fichiers `cgi` ne m'est pas accessible! Par contre, l'extension `pl` n'est pas proté-



Index of /DDG/documents/administrator/security

Name	Last modified	Size	Description
Parent Directory	22-Oct-2003 17:22	-	
WS_FTP.LOG	22-Oct-2003 13:46	1k	
groups	04-Oct-2001 00:00	1k	
passwords	04-Oct-2001 00:00	1k	

Apache/1.3.27 Server at linpc3.epfl.ch Port 80

la sécurité du site accessible depuis l'extérieur

Google EPFL intitle:"index of " cgi-bin

On peut être surpris qu'un contenu inaccessible depuis Google.ch soit accessible par le site de l'EPFL. La raison est simple, le ver est dans le fruit, simplement la boîte **jaune** connectée au réseau interne de l'EPFL lui permet d'indexer les contenus. Mais cette indexation entraîne sa visibilité depuis l'extérieur!

Les aficionados de la notion d'Intranet/Extranet vont encore s'arracher les cheveux...

Trous potentiels de sécurité, une attaque croisée

Un objectif des Google Hacks est de rechercher des vulnérabilités. Le côté magique réside dans leurs présentations sur un plateau. Par exemple un logiciel d'organisation de photos: `"Powered by Gallery" site:epfl.ch`. Bien, ce logiciel semble populaire, 56'600 pages uniquement dans l'EPFL!

Une brève recherche sur les failles de sécurité pour ce logiciel avec Google nous permet de découvrir une faille: ***http://www.securityfocus.com/bid/14668/info.***

Cette attaque utilise un moyen détourné: utilisant la capacité d'inclure des données descriptives supplémentaires dans une image, on inclut un code écrit en javascript. Ainsi, une action malveillante pourrait être encapsulée dans une simple image. Ce problème est connu depuis plus d'une année, et probablement les versions utilisées ne devraient plus être sensibles à cette vulnérabilité.

Scories révélatrices du passé

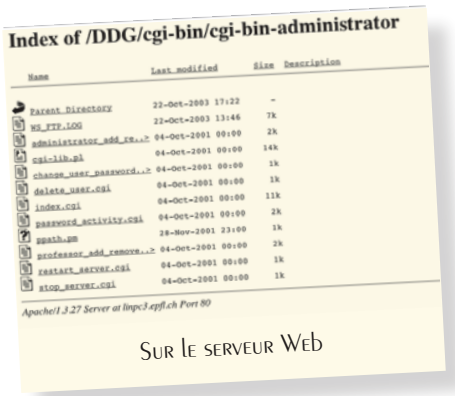
Le transfert de données entre un poste client et le serveur doit utiliser un protocole sécurisé.

Un programme largement répandu dans le monde Windows, WS_FTP, possède l'inconvénient de déposer dans le répertoire de destination, un fichier énumérant l'ensemble des modifications réalisées. Ce fichier fournit des informations importantes:

- la date de modification
- les répertoires et fichiers d'origine
- la machine, répertoire et fichier de destination.

La recherche est particulièrement simple, ***http://search.epfl.ch/ws_ftp.log***. Le côté amusant est que ce fichier suit les pérégrinations de l'hébergement du site, une fois créé, jamais supprimé!

On peut également vérifier les sites qui recommandent ce programme par la suppression de `"log"` dans les résultats. `"log" ws_ftp -log`.



Index of /DDG/cgi-bin/cgi-bin-administrator

Name	Last modified	Size	Description
Parent Directory	22-Oct-2003 17:22	-	
WS_FTP.LOG	22-Oct-2003 13:46	7k	
administrator_add_re..>	04-Oct-2001 00:00	2k	
cgi-lib.pl	04-Oct-2001 00:00	14k	
change_user_password..>	04-Oct-2001 00:00	1k	
delete_user.cgi	04-Oct-2001 00:00	1k	
index.cgi	04-Oct-2001 00:00	11k	
password_activity.cgi	04-Oct-2001 00:00	2k	
ppath.pm	28-Nov-2001 23:00	1k	
professor_add_remove..>	04-Oct-2001 00:00	2k	
restart_server.cgi	04-Oct-2001 00:00	1k	
stop_server.cgi	04-Oct-2001 00:00	1k	

Apache/1.3.27 Server at linpc3.epfl.ch Port 80

Sur le serveur Web

Google EPFL: "log" ws_ftp -log

Rechercher la première, ou la dernière photo

Si la recherche de vulnérabilité est l'objectif principal, on peut cependant utiliser le moteur de recherche pour des activités plus ludiques. Utiliser ces principes pour rechercher la première photo.

Les appareils photographiques SONY nomment automatiquement les photos prises par celui-ci avec le format: DSCxxxxx.jpg. La première photo est: DSC00001.jpg

Comme Google existe également pour la recherche d'image, voici la chaîne de recherche limitée aux serveurs domiciliés en Suisse: `DSC00001 site:.ch`. Pour la dernière une recherche par date dans l'ordinateur révèle que la dernière photo enregistrée est la 7717e, prise par mon fils de 5 ans!

Les remèdes

La première automédication est d'utiliser des Google Hack sur son propre site. Ainsi, les failles apparaissent immédiatement et elles peuvent être rapidement corrigées.

Par exemple, ne pas rendre public le contenu des dossiers sur un serveur Web, évitant l'utilisation du hack Google:

```
intitle:"index of" site:
epfl.ch
```

Sur un serveur Apache, la solution est simple:

```
<Directory /usr/local/
apache/htdocs>
Options -Indexes
</Directory.
```

Séparer clairement les contenus

Si une chaîne de hachage est conservée, elle ne doit jamais être située dans la hiérarchie du site Web, mais référencée par un chemin absolu sur le serveur.

On doit respecter la résolution qu'un contenu sécurisé doit être conservé dans un espace sécurisé, pas sur un site Web!

Ne pas se faire indexer

Tout contenu sensible ne doit pas être indexé, dans une logique absolue, il ne devrait même pas être visible, on tend vers le paradoxe du chat de Schrödinger, ***http://fr.wikipedia.org/wiki/Chat_de_Schrödinger.***

Pour un site

Créer un fichier *robots.txt* à la racine du site, ne pas oublier le s..., indiquer dans ce fichier les répertoires protégés contre l'indexation

```
User-agent: *
Disallow: /cgi-bin/
Disallow: /tmp/
Disallow: /private/
```

Pour une page Web

Utiliser le *meta tag* de l'indexation, à inclure dans l'en-tête de la page Web: `<meta name="robots" content="no index,nofollow">`.

Garder à jour l'environnement

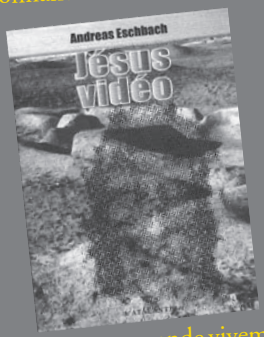
L'analyse des vulnérabilités des systèmes et outils est une activité de base d'un ingénieur système, particulièrement avec des ajouts non standards.

Dans le rôle de l'arroseur arrosé, il m'est arrivé la mésaventure d'héberger un site qui a été piraté, entraînant l'affichage modifié de la page de garde.

Par acquit de conscience, j'ai été sur le site répertoriant ce genre d'exploit, et surpris, ce site avait déjà été piraté sur deux autres serveurs. La chute de l'histoire est que c'était le responsable de ces serveurs qui m'avaient demandé d'héberger ce site! ■

Lectures d'été

Si on désire approfondir le sujet, Google m'oriente rapidement vers l'adresse du site du créateur de la notion de Google Hack: ***http://johnny.ihackstuff.com/.*** Le livre écrit par l'auteur du site présente une source d'inspiration quasi inépuisable. **Google Hacking for Penetration Testers, Johnny Long**, ISBN 1931836361, Syngress Media. Un élément important décrit dans cet ouvrage est la possibilité d'automatiser les recherches de failles de sécurité. Ainsi, la puissance de Google combinée avec un traitement informatique, offre un résultat impressionnant.



Je vous recommande vivement la lecture de **Jésus vidéo**, d'Andreas Eschbach, L'Atalante, ISBN 2-84172-167-1. L'auteur de science-fiction propose une vision autrement plus séduisante de la transmission des idéaux du christianisme que le livre à succès de Dan Brown, Da Vinci Code.